

Linear Response of General Observables in Spiking Neuronal Network Models.

Bruno Cessac¹ and Rodrigo Cofre²

¹Biovision team INRIA, Sophia Antipolis, France.

²Département de physique théorique, Université de Genève, Switzerland

April 19, 2017

Abstract

The activity of a neuronal network, characterized by action potentials (spikes) is constrained by the intrinsic properties of neurons and their interactions. When a neuronal network is submitted to external stimuli, the statistics of spikes changes, and becomes difficult to disentangle the influence of the stimuli from the intrinsic dynamics. Using the formalism of Gibbs distributions we analyze this problem in a specific model (Conductance-based Integrate-and-Fire), where the neuronal dynamics depends on the history of spikes of the network. We derive a linear response formula allowing to quantify the influence of a weak amplitude external stimuli on the average value of arbitrary observables. This formula clearly disentangles the effect of the stimuli, intrinsic neuronal dynamics, and network connectivity. Upon some approximations, it reduces to a convolution, allowing to recover a standard computation in computational neuroscience.

Keywords Neuronal Network Dynamics; Spike Train Statistics; Linear Response; Gibbs Distributions.

1 Introduction

Neurons communicate by sending short-lasting electrical signals called action potentials or “spikes”, which allow the rapid propagation of information throughout the brain. Novel experimental recording techniques permit to measure the collective spiking activity of larger and larger populations of interacting neurons responding to external stimuli [17]. These techniques allow, in particular, to better characterize the link between a stimuli and the responses of a neuronal network (referred in the literature as “the neural coding problem” [33]).

The main difficulty of characterizing this relationship lies in the fact that neuronal responses are highly variable [12]. It has been observed in a variety of experimental settings that, even at the single neuron level, when presenting repetitions of the same stimulus under controlled experimental conditions, the neural activity changes from trial to trial

[11, 39]; yet the statistics of the response is highly structured. Thus, researchers are seeking statistical regularities in order to unveil a probabilistic relation between stimuli and spiking responses [38, 42, 45].

Sensory neurons convey collectively to the brain information about external stimuli using correlated spike patterns [38, 42, 45]. This correlated firing has been linked to stimulus encoding [13], stimulus discrimination [2] and to intrinsic properties of the network, which remain even in absence of stimulus [44]. However, disentangling the biophysical origins of the correlations observed in spiking data is still a central and difficult problem in computational neuroscience [12, 18, 33], mainly because the mechanisms underlying correlated spiking are poorly understood. As a consequence, correlations in spiking neuronal networks have attracted a considerable amount of attention in the last years, from experimental data analysis perspectives [38, 28, 32, 45] and also from the theoretical modeling viewpoint [1, 41, 5, 43, 25, 29].

Mathematical models of spiking neuronal networks offer a complementary approach to biological experiments [21, 18]. These models describe the underlying dynamics of membrane potentials, evolving according to well known biophysical mechanisms [12, 18, 14] providing a framework to characterize the population spike train statistics in terms of parameters, representing the underlying synaptic connectivity, history of previous spikes and stimuli. In this paper we investigate, within this framework, the following question:

“How does a weak time-dependent stimuli change the spike train correlations for a given neuronal network dynamics and connectivity?”

There is a large body of theoretical work linking the spiking responses of neuronal network models to their structural properties in the presence of stimuli. One of the first papers along this line is [5] in which a network of homogeneous leaky integrate-and-fire neurons is considered in the mean-field limit. For constant external stimuli, the steady state solution of the Fokker-Planck equation characterizing the evolution of the distribution of the membrane potential allows finding the relation between stimuli and the firing rate (probability that a neuron spikes within a small time interval) as a function of the parameters of the model [18]. Finding the firing rate for time-varying stimuli is, however, a much more difficult problem. In [5, 25] using a different approach, i.e., analyzing the linear response in the frequency domain, a network of leaky integrate-and-fire neurons is considered and equations for cross-correlations are obtained. In both studies all neurons are considered identical and no precise connectivity structure is considered. More recently, two interesting works along the same line, but considering arbitrary network architecture have been published. In [43], the role of network connectivity is assessed. The authors obtain formulas for cross-correlations with an arbitrary delay in terms of “motifs” in the neuronal connectivity. In [31], the correlation structure in networks of interacting Hawkes processes is investigated. Both studies characterize correlations between pairs of neurons in terms of contributions of connectivity motifs.

All these works consider the statistics of simultaneous events: one neuron spike (when computing firing rates) or two neurons spike at the same time (when computing pairwise correlations). However, neuronal networks dynamics involves causal interactions between

neurons and strongly depends on the network history. As a consequence, the statistics of general events involving distinct neurons spiking at different times ought to be affected by stimuli and could shed light on the coding process. Obviously, a thorough characterization of such events gets rapidly out of reach in experimental data, as the number of neurons and time delays increases, because their probability is small and is not accurately measurable in experimental samples. We consider this question in neuronal networks models, using the appropriate theory.

We address this question based on ideas and methods from statistical mechanics. In particular, we use two standard tools in this field: Gibbs distributions and linear response theory. We consider a spiking neuronal network model whose spike train statistics is characterized by a Gibbs distribution. Obviously, the distribution changes if one applies an external stimuli. We analyze this change using linear response properties of the neuronal network model and its spike train statistics.

The linear response determines how the expectation value of an observable of a dynamical system changes upon changing a parameter of the dynamics. For stochastic systems, the linear response has been well known in the statistical mechanics community in the form of the fluctuation-dissipation theorem [23], which depends only on correlation functions with respect to the unperturbed system, meaning that the linear response can be predicted without perturbing the system. For deterministic systems, the theory of linear response has been well developed for chaotic dynamical systems. For hyperbolic dynamical systems, David Ruelle provided a proof of the linear response, as a consequence of the differentiability of the SRB measures, which describe the statistical properties of a physical system [35, 37]. However, it has recently been shown that for a large class of dynamical systems typical observables do not obey linear response as the invariant measure does not depend smoothly on the parameters [3].

This paper is organized as follows: In section 2, we introduce the spiking neuronal network model on which we focus our analysis [34]. We associate to it a discrete time stochastic process whose transition probabilities are explicitly written as a function of the parameters of the model and present the mathematical setting for the linear response analysis and Gibbs distributions. In section 3, we present our main result and compute the linear response in the context of perturbed Gibbs potentials for the model presented in section 2, written in terms of a correlation function and also provide the explicit formula based on a convolution kernel. In section 4 we discuss consequences and applications of our results, especially focus on population receptive fields. We finish this paper with discussions in section 5. For the sake of clarity, we have decided to focus our analysis on a particular spiking neuronal network model, which is a conductance-based Integrate and Fire model that we detail in the next section. However, our results are more general and can be applied to other spiking models. We have left the developments of the general result to the appendix.

2 The Spiking Neuronal Network Model

2.1 The Leaky Integrate-and-Fire model

Neuronal network models allows to understand how the parameters representing the structure of the neuronal network submitted to stimulus, shape the spiking response of the neuronal population and imprints the collective statistics of spikes. In this context, the Leaky Integrate-and-Fire model (LIF) is especially relevant because of its simplicity.

We recall here the definition of the LIF model and its extension the generalized Integrate-and-Fire (gIF) analyzed in this paper. We consider a formal neuron k , with membrane potential V_k , membrane capacity C_k , resistance R , submitted to a current $I_k(t)$. We define the firing threshold V_{th} such that:

$$C_k \frac{dV_k}{dt} + \frac{1}{R} V_k = I_k(t) \quad \text{if } V_k(t) < V_{th}, \quad (1)$$

(sub-threshold dynamics). If there is a time t_k such that the membrane potential of a given neuron k reaches the firing threshold, that is $V_k(t_k) \geq V_{th}$, the neuron k fires an action potential, i.e., it emits a spike and the membrane potential of neuron k is reset to a fixed reset value V_{res} instantaneously remaining at this value during a time denoted by Δ called “refractory period”, i.e., $V_k(t') = V_{res}$, $t' \in [t_k, t_k + \Delta]$. Equation (1) with the reset condition defines the LIF model introduced by Lapique in 1907 [24]. Despite its simplified nature, the LIF model captures some of the essential features of neuronal dynamics [18].

We assume that there is a minimal time scale denoted by δ , such that a neuron can at most fire one spike within a time interval of length δ . This provides a time discretization usually referred as “binning” labeled by an integer n . We denote the *spike-state* of each neuron $\omega_k(n) = 1$ whenever the k -th neuron emits a spike at discrete time n , i.e., $V_k(t) \geq V_{th}$ for some $n - 1 \leq t \leq n$, and $\omega_k(n) = 0$ otherwise, (see figure 1).

We consider a network of N neurons and denote by $\omega(n) := [\omega_k(n)]_{k=1}^N$ the spike-state of the entire network at time n , which we call *spiking pattern*. We note by $\mathcal{A} = \{0, 1\}^N$, the state space of spiking patterns in a network of N neurons; a *spike block* denoted by ω_m^n , $n \geq m$, is the sequence of spike patterns $\omega(m), \omega(m+1), \dots, \omega(n)$; blocks are elements of the set \mathcal{A}^{n-m+1} . The *time-range* (or “range”) of a block ω_m^n is $n - m + 1$, the number of time steps from m to n . The set of spike trains is $\Omega \equiv \mathcal{A}^{\mathbb{Z}}$.

2.2 Conductance Based Model

Many extensions of (1) have been proposed since 1907. Here, we consider a generalization introduced by Destexhe and Rudolph in [34], known as the gIF model, analyzed mathematically in [7, 9]. We consider this model, despite its complex dynamics depending on spike history, because it allows an analytic treatment giving access to the collective spike statistics. This is a key step toward the linear response characterization. In contrast to the LIF model (1), where the conductance was considered constant equal to $1/R$, here it depends on time and the spike history that we denote by ω . The stimulus also depends on time and spike history. We denote $g_k(t, \omega)$ the conductance, and $i_k(t, \omega)$ to the stimulus

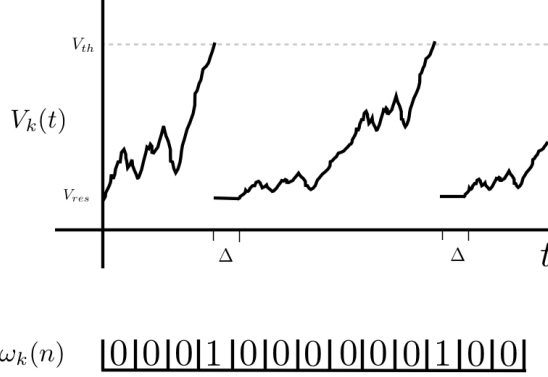


Figure 1: A sample of the time trajectory of the membrane potential of a given neuron k of equation (1) is plotted in continuous time. When the membrane potential reaches a fixed threshold V_{th} , is reset to a fixed value V_{res} and a spike is recorded in discrete time $\omega_k(n) = 1$, otherwise $\omega_k(n) = 0$.

of neuron k at time t depending both on the spike history up to time $\lfloor t \rfloor$. The gIF model reads,

$$C_k \frac{dV_k}{dt} + g_k(t, \omega) V_k = i_k(t, \omega) \quad \text{if } V_k(t) < V_{th}. \quad (2)$$

Upon firing of the pre-synaptic neuron j at (discrete) time $t_j(\omega)$, the membrane conductance of the post-synaptic neuron k is modified as [6]:

$$g_k(t, \omega) = \sum_{j=1}^N g_{kj}(t, \omega)$$

where

$$g_{kj}(t, \omega) = g_{kj}(t_j, \omega) + G_{kj} \alpha_{kj}(t - t_j), \quad t > t_j,$$

$G_{kj} \geq 0$ characterizes the maximal amplitude of the conductance during a post-synaptic potential. The function α_{kj} (called “alpha” profile) mimics the time course of the chemical synaptic conductance upon the occurrence of the spike. We assume that the alpha profiles are exponentially decaying functions:

$$\alpha_{kj}(t) = e^{\frac{-t}{\tau_{kj}}} H(t), \quad (3)$$

with $\alpha(0) = 0$, where τ_{kj} is the characteristic decay time and $H(t)$ is the Heaviside function. The current $I_k(t)$ in (1) is replaced by $i_k(t, \omega)$:

$$i_k(t, \omega) = g_{L,k} E_L + \sum_{j=1}^N W_{kj} \alpha_{kj}(t, \omega) + S_k(t) + \sigma_B \xi_k(t),$$

where, $g_{L,k}$ is the leak conductance, E_L is the leak reversal potential, W_{kj} is the synaptic weight from neuron j to neuron k , which is modulated by the alpha functions mimicking the synaptic decay (typically exponential) with time, $S_k(t)$ is the external stimulus received by neuron k , $\xi_k(t)$ is a white noise, and σ_B its amplitude.

We give a brief account of several essential results on the gIF model, already published in [7, 9].

2.3 Solution of the sub-threshold equation

As conductances do not depend explicitly on the membrane potential V_k (as e.g., in Hodgkin-Huxley model [20]) one can explicitly integrate the sub-threshold dynamics. For fixed¹ ω , equation (2) is a linear equation in V_k with flow $e^{-\frac{1}{C_k} \int_{t_1}^t g_k(u, \omega) du}$. This flow characterizes the membrane potential evolution of neuron k below threshold, i.e., when neuron k does not spike in the time interval $[t_1, t_2]$. One can easily extend the definition of the flow to include reset [9] by denoting $\tau_k(t, \omega)$ the last time before t where neuron k spikes in the spike-train ω , as follows:

$$\Gamma_k(t_1, t, \omega) = \begin{cases} e^{-\frac{1}{C_k} \int_{t_1}^t g_k(u, \omega) du} & \text{if } t \geq t_1 \geq \tau_k(t, \omega); \\ 0 & \text{otherwise.} \end{cases}$$

We can split the solution of (2) as follows:

$$V_k(t, \omega) = V_k^{(sp)}(t, \omega) + V_k^{(S)}(t, \omega) + V_k^{(noise)}(t, \omega). \quad (4)$$

where the first term on the r.h.s corresponds to the spontaneous part, which is independent of the external stimulus perturbation and noise. The spontaneous contribution can be divided again into a part corresponding to synapses and a part corresponding to the leak:

$$V_k^{(sp)}(t, \omega) = V_k^{(syn)}(t, \omega) + V_k^{(L)}(t, \omega), \quad (5)$$

$$V_k^{(syn)}(t, \omega) = \frac{1}{C_k} \sum_{j=1}^N W_{kj} \int_{\tau_k(t, \omega)}^t \Gamma_k(t_1, t, \omega) \alpha_{kj}(t_1, \omega) dt_1 \quad (6)$$

and;

$$V_k^{(L)}(t, \omega) = \frac{E_L}{\tau_{L,k}} \int_{\tau_k(t, \omega)}^t \Gamma_k(t_1, t, \omega) dt_1,$$

where:

$$\tau_{L,k} \stackrel{\text{def}}{=} \frac{C_k}{g_{L,k}}.$$

¹We assume that ω and the trajectory of membrane potentials are linked together by compatibility conditions (see [9]).

The second term in (4) corresponds to the contribution due to the external stimulus:

$$V_k^{(S)}(t, \omega) = \frac{1}{C_k} \int_{\tau_k(t, \omega)}^t S_k(t_1) \Gamma_k(t_1, t, \omega) dt_1. \quad (7)$$

The last one is the stochastic part of the membrane potential:

$$V_k^{(noise)}(t, \omega) = \frac{\sigma_B}{C_k} \int_{\tau_k(t, \omega)}^t \Gamma_k(t_1, t, \omega) dB_k(t_1),$$

where $B_k(t_1)$ is the standard Wiener process. We assume here for simplicity that the membrane potential is reset to 0 after spiking.

2.4 Transition Probabilities of the gIF Model

We now focus on the spike train statistics. The gIF model allows one to approximate in the limit of small σ_B the family of transition probabilities $\mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}]$ explicitly in terms of the parameters of the spiking neuronal network model. Note that we condition upon an infinite past, i.e., an event of zero probability, which defines a non-Markovian stochastic process known as “chain with complete connections” (see [7, 9] for details). Conditioning on an infinite past corresponds in fact to the following: The spike state of a neuron k at time n depends upon the spike history up to the last time in the past where k has fired, where the memory is reset. However, the time $\tau_k(n, \omega)$, is not bounded. This is mathematically well defined provided that specific conditions are fulfilled (see 2.7).

One can show that the gIF model as presented here² is conditionally independent [7].

$$\mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}] = \prod_{k=1}^N \mathbb{P} [\omega_k(n) \mid \omega_{-\infty}^{n-1}] \quad (8)$$

where,

$$\mathbb{P} [\omega_k(n) \mid \omega_{-\infty}^{n-1}] = \omega_k(n) \varphi(X_k(n-1, \omega)) + (1 - \omega_k(n)) (1 - \varphi(X_k(n-1, \omega))), \quad (9)$$

and

$$X_k(n-1, \omega) = \frac{\theta - V_k^{(det)}(n-1, \omega)}{\sigma_k(n-1, \omega)}, \quad (10)$$

where $V_k^{(det)}$ is the deterministic solution of the membrane potential at time $(n-1)$.

$$V_k^{(det)} \equiv V_k^{(sp)} + V_k^{(S)}, \quad (11)$$

(see (4),(5),(7)),

$$\sigma_k^2(n, \omega) = \left(\frac{\sigma_B}{C_k} \right)^2 \int_{\tau_k(t, \omega)}^n \Gamma_k^2(t_1, n, \omega) dt_1,$$

²A more complete version if this model also includes electric synapses [9], in that case the conditional independence is lost.

and

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{u^2}{2}} du.$$

There exists a relationship between discrete-time stochastic processes (chains) and one-dimensional Gibbs measures which has been characterized mathematically in [16]. While discrete-time stochastic processes are defined by transition probabilities, Gibbs measures are defined in terms of specifications, which determine its finite-volume conditional probabilities when the exterior of the volume is known. In one “time” dimension this is equivalent to conditioning both the past and the future. For stochastic processes with exponential continuity rate, it has been shown in [16] that a set of transition probabilities are equivalent to the so-called *left-interval specifications* (LIS), which characterize a unique Gibbs measure. The exponential continuity rate is satisfied by the transition probabilities of the gIF model [7, 9], therefore the family of transition probabilities (8) characterizes uniquely a Gibbs distribution.

2.5 Observables and Monomials

We call *observable* a function that associates a real number to a spike-train. We say that the observable $f : \Omega \rightarrow \mathbb{R}$ has range $R = D + 1$ if $f(\omega) \equiv f(\omega_0^D)$, i.e., f depends only on the first R time steps in the spike-train. Examples of observables are the *monomials*, i.e., observables of the form:

$$m_l(\omega) = \prod_{k=1}^L \omega_{i_k}(t_k).$$

One fixes a set of pairs $\{(i_k, t_k)\}_{k=1}^L = l$ (neuron index, time index) and $m_l(\omega) = 1$ if and only if neuron i_k spikes at time $t_k, \forall k$ in the spike-train ω , and $m_l(\omega) = 0$ otherwise. In a range R monomial, the firing times t_k are constrained within the interval $\{0, \dots, D\}$ with $D \geq 0$. There are finitely many such possible monomials and one can index each of them by an index l in one-to-one correspondence with the set of pairs (i_k, t_k) . Any range- R observable can be written as a linear combination of range- R monomials (for details we refer to [10]). For a range R -observable f we write its monomial decomposition as follows:

$$f(\omega) = \sum_l f_l m_l(\omega), \quad (12)$$

where $f_l \in \mathbb{R}$ is thus the coefficient of monomial m_l in the decomposition.

Note that here, we also consider observables depending explicitly on time, denoted $f(t, \omega)$ considering that observables are causal. When we write $f(t, \omega)$ we mean that f depends on spikes occurring *before* time t , i.e., if we note $[t] = n$, where $[t]$ is the largest integer smaller than t , $f(t, \omega)$ stands for $f(t, \omega_{-\infty}^n)$. For a range R observable the monomial decomposition (12) becomes:

$$f(t, \omega) = \sum_l f_l(t) m_l(\omega_{n-D}^n).$$

2.6 Schematic representation

In the previous sections, we have presented the gIF spiking neuronal network model. We have shown some results which allows the explicit characterization of the spike train statistics in terms of the parameters of the model. Clearly, the statistics will change if one applies an external stimulus. Consider the gIF model evolving from time $t = -\infty$ until time $t = t_0$ without external stimuli, that is, $S_r(t) = 0, \forall r, \forall t \leq t_0$ in (2). From time t_0 a weak amplitude time-dependent stimulus is applied. Our main goal is to analyze how the statistics of a given observable changes after the stimuli is applied as a function of the connectivity structure and dynamics of the spiking neuronal network model (see figure 2).

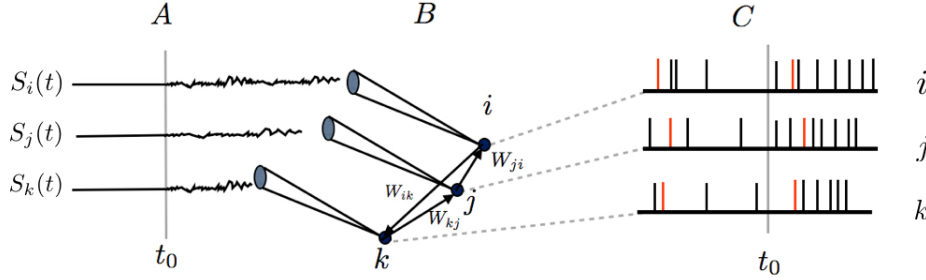


Figure 2: *A*: The weak time-dependent stimuli are switched on at time t_0 . *B*: A network of interconnected neurons (here only composed by neurons i, j and k) is spiking spontaneously before t_0 , after it is responding to time-dependent stimuli. *C*: The spike trains of neurons i, j and k . In this example we focus on the following observable (in red): “neurons i and k firing one time step before neuron j ”, which has some statistics before and after the stimuli is applied. In this paper we are interested in characterizing the change in average of any observable after t_0 , as a function of the parameters of the model.

2.7 Gibbs Distributions from Transition Probabilities

In this section, we present the mathematical framework we use to characterize the change in the statistics of observables. We start presenting how to characterize a Gibbs distribution consistent with a given family of transition probabilities or (LIS) and the average values of observables over spike trains with respect to the underlying Gibbs distribution.

A *Gibbs distribution* is a probability measure μ such that:

(A) For all $n \in \mathbb{Z}$ and all $\mathcal{F}_{\leq n}$ -measurable functions f^3 :

$$\int f(\omega_{-\infty}^n) \mu(d\omega) = \int \sum_{\omega(n) \in \mathcal{A}} f(\omega_{-\infty}^{n-1} \omega(n)) \mathbb{P}[\omega(n) \mid \omega_{-\infty}^{n-1}] \mu(d\omega).$$

³Measurable with respect to the natural filtration associated to its “past” at each time

(B) For all $n \in \mathbb{Z}$, $\forall \omega_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$, $\mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}] > 0$.

(C) For each $n \in \mathbb{Z}$, $\mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}]$ is continuous with respect to ω^4 .

The condition (A) is a natural extension (for infinite memory processes) of the condition defining the invariant probability of an homogeneous Markov chain. Existence and uniqueness of a Gibbs distribution is guaranteed by (B) and (C). We have shown in [9] that the transition probabilities of the gIF model satisfy the conditions to have a unique Gibbs distribution, in the presence of a (time-dependent) stimulus. Moreover,

$$\mu [\omega_m^n] = \int_{\mathcal{A}_{-\infty}^{m-1}} \mathbb{P} [\omega_m^n \mid \omega_{-\infty}^{m-1}] \mu(d\omega), \quad (13)$$

By definition, the expectation of $f(t, \omega)$ with respect to μ is:

$$\mu [f(t, \cdot)] = \int f(t, \omega) \mu(d\omega) = \sum_{\omega(n) \in \mathcal{A}} \int f(t, \omega_{-\infty}^{n-1} \omega(n)) \mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}] \mu(d\omega), \quad (14)$$

where the last equality comes from condition (A) and where $n = \lceil t \rceil$. If f is a range R observable we have:

$$\mu [f(t, \cdot)] = \sum_{\omega_{n-D}^n} f(t, \omega_{n-D}^n) \mu [\omega_{n-D}^n],$$

Note that, using the decomposition (12), we can decompose the average $\mu [f(t, \cdot)]$ as:

$$\mu [f(t, \cdot)] = \sum_{\omega_{n-D}^n} \sum_l f_l(t) \mu [m_l(\omega_{n-D}^n)],$$

where the average $\mu [m_l(\omega_{n-D}^n)]$ depends on n as μ is non stationary in general.

2.8 Normalized Gibbs Potential

An alternative approach to characterize Gibbs distributions is based on potentials, which has been rather successful in statistical physics, this is the approach we use here. We define the *normalized Gibbs potential* $\phi(n, \omega) : \mathbb{N} \times \Omega \rightarrow \mathbb{R}$ by:

$$\phi(n, \omega) \equiv \log (\mathbb{P} [\omega(n) \mid \omega_{-\infty}^{n-1}]) \quad (15)$$

and

$$\mathbb{P} [\omega_m^n \mid \omega_{-\infty}^{m-1}] = e^{\Phi(m, n, \omega)}, \quad (16)$$

with:

$$\Phi(m, n, \omega) = \sum_{l=m}^n \phi(l, \omega).$$

⁴The dependence with respect to the past spikes decay exponentially fast with time.

Then,

$$\mu[\omega_m^n] = \int_{\mathcal{A}_{-\infty}^{m-1}} e^{\Phi(m,n,\omega)} \mu(d\omega).$$

Following the definition (15), we obtain the normalized potential (9) for the gIF model.

$$\phi(n, \omega) = \sum_{k=1}^N \phi_k(n, \omega), \quad (17)$$

where

$$\phi_k(n, \omega) = \omega_k(n) \log \varphi(X_k(n-1, \omega)) + (1 - \omega_k(n)) \log(1 - \varphi(X_k(n-1, \omega))), \quad (18)$$

which depends on integer time, via the variable $X_k(n-1, \omega)$ (10).

3 Linear Response of the gIF Model

3.1 First order variation

The explicit form of the normalized Gibbs potential computed in the previous section allows us to apply the linear response theory in the classical “perturbative” framework of the so-called linear response theory in non-equilibrium statistical physics [22, 36]. This normalized potential has the property that can be separated into a “spontaneous” part $\phi^{(sp)}(\omega)$, which is independent of the stimuli and time and a “perturbation” part $\delta\phi(n, \omega)$ depending on a weak time-dependent stimuli, which is non-zero from time t_0 . Following (17) we write:

$$\phi(n, \omega) = \begin{cases} \phi^{(sp)}(\omega) & \text{if } n < [t_0]; \\ \phi^{(sp)}(\omega) + \delta\phi(n, \omega) & \text{if } n \geq [t_0]. \end{cases} \quad (19)$$

We note $\mu^{(sp)}$ the Gibbs distribution characterizing the spike train statistics before the application of the external stimulus, i.e., during the time interval $] -\infty, t_0]$. We assume that $\mu^{(sp)}$ is a stationary measure. We denote μ the Gibbs distribution in the presence of the external stimuli. It characterizes the spike train statistics from time t_0 , thus is in general *non-stationary*, as stimuli depend on time. Note that, if the external stimuli is switched on at time t_0 , spike statistics is still constrained by the previous spontaneous activity, since transition probabilities have memory. This effect is especially salient in the gIF model which has an infinite memory.

We define the *first order variation of the observable* $f(t, \omega)$ in the presence of a stimuli denoted $\delta^{(1)}\mu[f(t, \cdot)]$:

$$\delta^{(1)}\mu[f(t, \cdot)] = \mu[f(t, \cdot)] - \mu^{(sp)}[f(t, \cdot)],$$

If $t \leq t_0$, $\delta^{(1)}\mu[f(t, \cdot)] = 0, \forall f(t, \cdot)$ as $\mu = \mu^{(sp)}$. We define the *correlation* between two observables $f(t, \omega)$ and $g(r, \omega)$ under $\mu^{(sp)}$, as follows;

$$\mathcal{C}^{(sp)}[f(t, \cdot), g(r, \cdot)] = \mu^{(sp)}[f(t, \cdot)g(r, \cdot)] - \mu^{(sp)}[f(t, \cdot)]\mu^{(sp)}[g(r, \cdot)],$$

We consider $n_0 = \lceil t_0 \rceil$, and denote

$$\delta\Phi(n_0 + 1, n, \omega) = \sum_{r=n_0+1}^{n=\lceil t \rceil} \delta\phi(r, \omega), \quad (20)$$

for n integer.

With these definitions we formally obtain the first order variation of the observable $f(t, \omega)$ which is written as a correlation between the observable $f(t, \cdot)$ and the integrated perturbation potential function $\delta\Phi$. (See appendix A):

$$\begin{aligned} \delta^{(1)}\mu[f(t, \cdot)] &= \mathcal{C}^{(sp)}[f(t, \cdot), \delta\Phi(n_0 + 1, n, \cdot)] \\ &= \sum_{r=n_0+1}^{n=\lceil t \rceil} \mathcal{C}^{(sp)}[f(t, \cdot), \delta\phi(r, \cdot)]. \end{aligned} \quad (21)$$

Comments

- Eq. (21) is a version of the so-called *fluctuation-dissipation theorem* in statistical physics [22, 36]. Here, it holds for Gibbs distributions with infinite range potential $\phi(t, \omega)$. As stated in [8] a key assumption is that $\phi(t, \omega)$ is continuous with respect to ω with a modulus of continuity decaying sufficiently fast (e.g. exponentially). Eq. (21) applies to fairly general models under the assumption that a unique Gibbs distribution for spike trains exists. In the gIF model, this condition is ensured by the exponential tail in synaptic responses, which in our model is due to the alpha functions (3), a fairly general property of chemical synapses [7, 8]. We expect therefore (21) to hold for a wide range of spiking neuronal network models.
- One of the advantages of this relation is that averages are taken with respect to $\mu^{(sp)}$, which can be approximated by *empirical averages* on spontaneous activity, for example, from experimental data using the maximum entropy principle.

3.2 Spontaneous and Perturbed Gibbs Potential of the gIF model

In this section, we split explicitly the Gibbs potential of the gIF model (18) in the spontaneous part $\phi^{(sp)}(\omega)$ and the one evoked by the stimuli $\delta\phi(n, \omega)$ as in (19). This allows us to apply the linear response on general observables (21).

Switching on the stimulus introduces a variation of $V_k^{(det)}$ (eq. (11)), which in turn causes a variation of X_k (eq. (10)). This function can be written as follows:

$$X_k(n, \omega) = X_k^{(sp)}(n, \omega) + \delta X_k(n, \omega), \quad (22)$$

where $X_k^{(sp)}(n, \omega)$ is independent of the stimuli. From (7), (10):

$$\delta X_k(n, \omega) = -\frac{1}{C_k \sigma_k(n, \omega)} \int_{\tau_k(n, \omega)}^n S_k(t_1) \Gamma_k(t_1, n, \omega) dt_1,$$

for $t \geq 0$. From (18) we have for integer time r :

$$\begin{aligned} \phi_k(r, \omega) &= \omega_k(r) \log \varphi \left(X_k^{(sp)}(r-1, \omega) + \delta X_k(r-1, \omega) \right) \\ &\quad + (1 - \omega_k(r)) \log \left(1 - \varphi \left(X_k^{(sp)}(r-1, \omega) + \delta X_k(r-1, \omega) \right) \right). \end{aligned}$$

In the remainder of this section we write $X_k^{(sp)}, \delta X_k$ instead of $X_k^{(sp)}(r-1, \omega), \delta X_k(r-1, \omega)$ to alleviate notations.

We make a series expansion of $\log \varphi(X_k^{(sp)} + \delta X_k)$ at $X_k^{(sp)}$.

$$\begin{aligned} \log \varphi(X_k^{(sp)} + \delta X_k) &= \log \varphi(X_k^{(sp)}) + \sum_{u=1}^{+\infty} \frac{a_u(X_k^{(sp)})}{u!} (\delta X_k)^u, \\ \log \left(1 - \varphi(X_k^{(sp)} + \delta X_k) \right) &= \log \left(1 - \varphi(X_k^{(sp)}) \right) + \sum_{u=1}^{+\infty} \frac{b_u(X_k^{(sp)})}{u!} (\delta X_k)^u \end{aligned}$$

where a_u and b_u are the u -th derivative of $\log \varphi(x)$ and $\log(1 - \varphi(x))$ respectively:

$$a_1(x) = \frac{\varphi'(x)}{\varphi(x)} = -\frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{x^2}{2}}}{\varphi(x)}, \quad b_1(x) = -\frac{\varphi'(x)}{1 - \varphi(x)},$$

Therefore,

$$\phi_k(r, \omega) = \phi_k^{(sp)}(r, \omega) + \delta \phi_k(r, \omega),$$

with a time-dependent perturbation

$$\delta \phi_k(r, \omega) = \sum_{u=1}^{+\infty} \delta \phi_k^{(u)}(r, \omega),$$

where:

$$\delta \phi_k^{(u)}(r, \omega) = \mathcal{H}_k^{(u)}(r, \omega) (\delta X_k(r-1, \omega))^u,$$

with:

$$\mathcal{H}_k^{(u)}(r, \omega) \stackrel{\text{def}}{=} \frac{1}{u!} \left(\omega_k(r) a_u \left(X_k^{(sp)}(r-1, \omega) \right) + (1 - \omega_k(r)) b_u \left(X_k^{(sp)}(r-1, \omega) \right) \right).$$

We have then,

$$\begin{aligned}
\delta X_k(r-1, \omega) &= -\frac{1}{C_k \sigma_k(r-1, \omega)} \int_{\tau_k(r-1, \omega)}^{r-1} S_k(t_1) \Gamma_k(t_1, r-1, \omega) dt_1 \\
&= -\frac{1}{C_k \sigma_k(r-1, \omega)} \int_{\max(t_0, \tau_k(r-1, \omega))}^{r-1} S_k(t_1) \Gamma_k(t_1, r-1, \omega) dt_1.
\end{aligned}$$

where the last equality holds because $S(t_1) = 0$ for $t_1 < t_0$. Up to second order we finally obtain:

$$\delta \phi_k^{(1)}(r, \omega) = -\frac{\mathcal{H}_k^{(1)}(r, \omega)}{C_k \sigma_k(r-1, \omega)} \int_{\max(t_0, \tau_k(r-1, \omega))}^{r-1} S_k(t_1) \Gamma_k(t_1, r-1, \omega) dt_1,$$

The function $\delta \phi_k^{(1)}(r, \omega)$ is the first order variation of normalized potential when neurons are submitted to a weak time dependent stimulus, under the approximation (18). This is a function of the parameters defining the gIF neuronal network model.

3.3 Linear response of a general observable

From equation (21) we obtain the linear response of the general observable $f(t)$:

$$\delta^{(1)} \mu[f(t)] = - \sum_{r=n_0+1}^{n=\lceil t \rceil} \mathcal{C}^{(sp)} \left[f(t, \cdot), \sum_{k=1}^N \frac{\mathcal{H}_k^{(1)}(r, \cdot)}{C_k \sigma_k(r-1, \cdot)} \int_{\max(t_0, \tau_k(r-1, \cdot))}^{r-1} S_k(t_1) \Gamma_k(t_1, r-1, \cdot) dt_1 \right]. \quad (23)$$

All functions with a dot are random functions (via ω) whose law is constrained by the spontaneous Gibbs distribution $\mu^{(sp)}$. This is the case for the last firing time $\tau_k(r-1, \cdot)$, $\mathcal{H}_k^{(u)}(r, \cdot)$ and $\sigma_k(r-1, \cdot)$. These last two quantities are defined via time-integrals where the lower bound is $\tau_k(r-1, \cdot)$

$$\sigma_k^2(r-1, \omega) = \left(\frac{\sigma_B}{C_k} \right)^2 \int_{\tau_k(r-1, \omega)}^{r-1} \Gamma_k^2(t_1, r-1, \omega) dt_1,$$

$$\mathcal{H}_k^{(1)}(r, \omega) = \omega_k(r) a_1 \left(X_k^{(sp)}(r-1, \omega) \right) + (1 - \omega_k(r)) b_1 \left(X_k^{(sp)}(r-1, \omega) \right).$$

This is the most general expression we obtain for this model. It is valid for any observable, thus even for long range (spatio-temporal) correlations.

3.3.1 Role of synaptic connectivity in the linear response

Equation (23) carries the information about the connectivity structure through $\mathcal{H}_k^{(1)}(r, \omega)$ which is a function of $X_k^{(sp)}(r-1, \omega)$ (22). This function depends on the spontaneous part of the membrane potential $V_k^{(sp)}$ (5), which in turn, is a function of $V_k^{(syn)}$ (6), which contains the information about the synaptic connectivity structure of the network through the synaptic weights W_{ij} .

3.4 Linear Response of Monomials

Our result is valid for general observables, however the main application to spike train statistics requires the arguments in the correlation function (21) to be written in terms of monomials (such as firing rates, synchronous pairwise correlations and other spatio-temporal correlations), which can always be done using (12).

Assuming that the infinite range potential (15) can be approximated by a range- R potential (see [15, 10]) and using the decomposition (12), we have:

$$\delta\phi(r, \omega) \sim \sum_l \delta\phi_l(r) m_l(\omega_{r-D}^r).$$

Let us also assume that f has range R too; so that $f(t, \omega) = \sum_l f_l(t) m_l(\omega_{n-D}^n)$. We introduce the centered observable $\bar{f}(t) = f(t) - \mu^{(sp)}[f(t)]$ (and similarly $\delta\phi(r) = \delta\phi(r) - \mu^{(sp)}[\delta\phi(r)]$). We have $\mu^{(sp)}[f(t, \cdot)] = \sum_l f_l(t) \mu^{(sp)}[m_l]$, so that $\bar{f}(t) = \sum_l f_l(t) \bar{m}_l(\omega_{n-D}^n)$ with $\bar{m}_l = m_l - \mu^{(sp)}[m_l]$ and $n = [t]$. In this setting, (21) becomes:

$$\begin{aligned} \delta^{(1)}\mu[f(t)] &= \sum_{r=n_0+1}^{n=[t]} \mathcal{C}^{(sp)}[f(t, \cdot), \delta\phi(r, \cdot)] \\ &= \sum_{r=r_0+1}^{n=[t]} \sum_{l, l'} f_l(t) \delta\phi_{l'}(r) \mathcal{C}^{(sp)}[\bar{m}_l(\omega_{n-D}^n), \bar{m}_{l'}(\omega_{r-D}^r)]. \end{aligned}$$

If the observable $f(t)$ is a single monomial $f_l(t)$, the previous equation simplifies:

$$\delta^{(1)}\mu[f_l(t)] = \sum_{r=n_0+1}^{n=[t]} \sum_{l'} f_l(t) \delta\phi_{l'}(r) \mathcal{C}^{(sp)}[\bar{m}_l(\omega_{n-D}^n), \bar{m}_{l'}(\omega_{r-D}^r)]. \quad (24)$$

A similar approach (decomposition of observables on a suitable basis) has been used by Lucarini et al in a different context [26].

When $f(t, \omega)$ is a finite range and time independent monomial : $f(t, \omega) = m_l(\omega_{n-D}^n)$ and the function $\delta\phi(r, \omega) = \delta\phi(r, \omega_{n_0}^r)$ is also a finite range observable, this last one can be represented in the monomial basis, i.e., it can be written $\delta\phi(r, \omega_{n_0}^r) = \sum_{l'} \delta\phi_{l'}(r) m_{l'}(\omega_{n_0}^r)$. Thus:

$$\delta^{(1)}\mu[m_l(\omega)] = - \sum_{r=n_0+1}^{n=[t]} \sum_{l'} \delta\phi_{l'}(r) \mathcal{C}^{(sp)}[m_l(\omega), m_{l'}(\omega_{n_0}^r)]. \quad (25)$$

This representation allows one to obtain the linear response of a spatio-temporal observable $m_l(\omega)$ by computing correlations between monomials with respect to the spontaneous measure.

Remark: For the gIF model both the stimulus and the connectivity structure are present in equation (25) in the terms $\delta\phi_{l'}(r)$. Therefore, independent of the monomial that we consider to measure, here denoted $m_l(\omega)$ **we cannot ignore the connectivity architecture of the network**.

3.5 Linear Response as a Convolution

We consider the following notation: $C_{l,l'}$ stands for $\mathcal{C}^{(sp)}[\bar{m}_l, \bar{m}_{l'}]$ and $C_{l,l'}(n-r)$ for the correlation between the monomial \bar{m}_l and $\bar{m}_{l'}$ shifted $(n-r)$ time steps. From equation 24 and considering the translation invariance of $\mu^{(sp)}$ we obtain the following:

$$\delta^{(1)}\mu[f(t)] = \sum_{r=-\infty}^{n=[t]} \sum_{l,l'} f_l(t) C_{l,l'}(n-r) \delta\phi_{l'}(r), \quad (26)$$

This formula can be extended to the case where f has an arbitrary range, but the main interest is that linear response can be written in terms of correlations between range R monomials.

The linear response can also be written in the classical form of convolution. Indeed, introducing the discrete time convolution:

$$(g * h)[n] \equiv \sum_{r=-\infty}^n g(r)h(n-r),$$

we can write (26) in the form:

$$\delta^{(1)}\mu[f(t)] = \sum_l f_l(t) \sum_{l'} (\delta\phi_{l'} * C_{l,l'})[n].$$

Finally, defining the L dimensional vector $(\delta\phi * C)[n]$ with entries $(\delta\phi * C)[n]_l \equiv \sum_{l'} (\delta\phi_{l'} * C_{l,l'})[n]$, and the Euclidean scalar product on \mathbb{R}^L , $\langle u | v \rangle = \sum_l u_l v_l$ we obtain:

$$\delta^{(1)}\mu[f(t)] = \langle f(t) | (\delta\phi * C)[n] \rangle, \quad n = [t].$$

Remark: Here the convolution is between the parameters associated to the monomial decomposition of the perturbed potential (which contain the information about the stimuli, connectivity and other parameters representing biophysical properties of the network) and the correlation function between monomials of the perturbed potential and the monomials associated to the observable $f(t)$.

4 Consequences

In computational neuroscience, it is common to characterize the spiking activity of a neuron by its firing rate $r(t)$, which represent the probability that this neuron spikes during a small interval $[t, t+dt]$. Under the influence of an external stimulus the firing rate changes.

A classical ansatz, coming from the more general Volterra expansion (see [33] for an introduction within this field) is to write the firing rate of a neuron as a convolution, when the stimulus S is weak enough.

$$\delta^{(1)}[r(t)] = (K * S)[t] \quad (27)$$

where K is a convolution kernel depending on neuronal properties. In particular, for sensory neurons, when S is an external stimulus (e.g., a sequence of images presented to a retinal ganglion neuron), K is called the “receptive field” as it characterizes how a neuron respond to a region of space with an elementary stimulus (Dirac distribution or white noise). Although receptive fields can be computed from experimental data, the equation (27) has been mathematically justified only in a few models [18, 30, 32].

Here, we would like to show that equation (27) holds only under specific assumptions on the gIF model. In order to obtain a convolution between a “kernel” and the external stimulus from equation (23), we need to be able to take the integral and external stimulus S out of the correlation function. In order to do this, we have to consider that the lower limit of the definite integral inside the correlation in (23) is a function of the spontaneous measure, under which the correlations are computed. However, in the limit $\max(t_0, \tau_k(r - 1, \cdot)) \rightarrow -\infty$, it becomes independent of the spontaneous measure and deterministic. On the other hand, the external stimulus is a deterministic function which can be taken out of the correlation function. Under this assumption we obtain:

$$\delta^{(1)}\mu[f(t)] = -\sum_{k=1}^N \frac{1}{C_k} \int_{-\infty}^t \sum_{r=-\infty}^{n=[t]} \mathcal{C}^{(sp)} \left[f(t, \cdot), \frac{\mathcal{H}_k^{(1)}(r-1, \cdot)}{\sigma_k(r-1, \cdot)} \Gamma_k(t_1, r-1, \cdot) \right] S_k(t_1) dt_1.$$

As correlations are computed with respect to $\mu^{(sp)}$ which is time-translation invariant we have:

$$\begin{aligned} & \mathcal{C}^{(sp)} \left[f(t, \cdot), \frac{\mathcal{H}_k^{(1)}(r, \cdot)}{\sigma_k(r-1, \cdot)} \Gamma_k(t_1, r-1, \cdot) \right] \\ &= \mathcal{C}^{(sp)} \left[f(t-t_1, \cdot), \frac{\mathcal{H}_k^{(1)}(r', \cdot)}{\sigma_k(r'-1, \cdot)} \Gamma_k(0, r'-1, \cdot) \right], \end{aligned}$$

where $r' = [r - t_1]$. Denoting:

$$\kappa_{k,f}(t-t_1) = \frac{1}{C_k} \sum_{r=-\infty}^{[t-t_1]} \mathcal{C}^{(sp)} \left[f(t-t_1, \cdot), \frac{\mathcal{H}_k^{(1)}(r, \cdot)}{\sigma_k(r-1, \cdot)} \Gamma_k(0, r-1, \cdot) \right], \quad (28)$$

we obtain:

$$\delta^{(1)}\mu[f(t)] = \sum_{k=1}^N \int_{-\infty}^t \kappa_{k,f}(t-t_1) S_k(t_1) dt_1$$

which can be written as a convolution:

$$\delta^{(1)}\mu[f(t)] = (\kappa_f * S)[t],$$

where (28) is a vector with entries $\kappa_{k,f}$, which contains the model-dependence.

Remark: The condition $\max(t_0, \tau_k(r-1, \cdot)) \rightarrow -\infty$ under which we can take out the integral of the correlation is very strong and does not hold in general. This situation corresponds to spiking neuronal network which spike rarely and where the correlations are computed long time after the presentation of the stimulus. However, experimental data from a population of 160 retinal ganglion cells of salamander submitted to natural stimuli spike rarely [42], in particular there are sub-populations of neurons that fire very scarcely for which this assumption is not far from real.

5 Discussion

We formally obtained the linear response function in terms of the parameters of a spiking neuronal network model and the spike history of the network. The novelty of our approach is that it provides a consistent treatment of the expected perturbation of higher-order interactions, going in this way beyond the known linear perturbation of firing rates and instantaneous pairwise correlations, in particular extends to time-dependent correlations. Our linear response formula is written in terms of a correlation function with respect to the unperturbed measure as in the fluctuation-dissipation theorem in statistical physics [23, 36], which holds for Gibbs distributions with infinite range potential.

The explicit form of the Gibbs potential (18) in terms of the parameters of the gIF model allows the application of linear response techniques in the perturbative framework of time-dependent statistical mechanics of classical systems. The Gibbs framework considering infinite memory and time-dependent potentials is particularly useful for this purpose, as it describes the correlations in terms of structure and dynamics of the model [6, 7, 9, 10].

Although a linear treatment may seem a strong simplification, our results suggest that the connectivity architecture should not be neglected, in comparison to eq. (27). For the gIF model, we showed explicitly how the structure of synaptic connectivity appear in equation (25), that means that the perturbed average of a general observable is not only due to the interaction of this with the stimuli. In the presence of stimuli, the whole architecture of synaptic connectivity, history and dynamical properties of the networks are playing a role in the correlations through the perturbed potential. This agrees well with results from a recent study exhibiting an exact analytical mapping between neuronal network models and maximum-entropy models, showing that in order to accurately describe the statistical behavior of any observable in the maximum entropy model all the synaptic weights are needed, even to predict firing rates of single neurons [10]. Our results extend beyond the gIF model and can be directly applied to other spiking neuronal network models from which the normalized Gibbs potential can be computed. For example, the Generalized Linear Models [8], the discrete-time integrate-and-fire model [6], or the gIF model including electrical synapses [9]. For these models, the potential can be easily split in the same way as done in section 3.2 for the gIF model, and equation (21) can be used to predict the linear response of an arbitrary observable.

In sensory networks, as the mammalian retina, the network architecture together with dynamical properties of the network, influence the statistics of spike correlations produced by the neuronal network in order to perform the collective operation of conveying information about the visual stimuli to different areas in the visual cortex [4]. Therefore, it seems

plausible that for repetitions of a given time-dependent stimuli, if the synaptic connectivity between different layers of the retina change (due to pharmacological intervention or plasticity), the correlations between spiking events and the firing rates change as well [40]. In particular, we have shown that our results help to provide a more general characterization of receptive fields, including the connectivity structure in which the neurons are embedded even at the single neuron level.

Another possible application of our result come from the fact that correlations are taken with respect to the spontaneous measure, which can be approximated from experimental recordings of a neuronal network spiking in absence of external stimuli using the maximum entropy principle. As correlations between monomials can also be computed from data, assuming that the neuronal tissue from which the spikes have been recorded can be modeled by the gIF, only the values of the parameters of the model are needed to compute $\delta\phi$ and to predict the linear response.

In a time where considerable efforts are dedicated to the construction of detailed connection maps of neurons on multiple scales [19, 27], the values of these parameters will be more and more accurate and available to theoreticians willing to test their predictions. We believe that the next few years may be crucial for the understanding of how network structure influences neuronal correlation.

APPENDIX

A Linear Response of Spiking Neuronal Network Models

Suppose we can decompose the normalized potential $\phi(t, \omega) = \phi^{(sp)}(\omega) + \delta\phi(t, \omega)$, where $\phi^{(sp)}(\omega)$ does not depend on time, with $\delta\phi(t, \omega) = 0$ for $t < t_0$. As a result of this weak time-dependent “external field”, there is an additional time-dependent term in the Gibbs potential.

From the definition (16), $e^{\Phi^{(sp)}}$ corresponds to the family of transition probabilities $\mathbb{P}^{(sp)}$ defining the Gibbs distribution $\mu^{(sp)}$ in the spontaneous regime, whereas e^{Φ} corresponds to the family of transition probabilities \mathbb{P} defining the Gibbs distribution μ in time-dependent stimuli-evoked regime. For $n > n_0$ we have:

$$e^{\Phi(n_0+1, n, \omega)} = e^{\Phi^{(sp)}(n_0+1, n, \omega) + \delta\Phi(n_0+1, n, \omega)} = e^{\Phi^{(sp)}(n_0+1, n, \omega)} \left[1 + \sum_{p=1}^{+\infty} \frac{\delta\Phi(n_0+1, n, \omega)^p}{p!} \right],$$

which from equation (16) gives:

$$\mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] = \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \left[1 + \sum_{p=1}^{+\infty} \frac{\delta\Phi(n_0+1, n, \omega)^p}{p!} \right].$$

Taking the first order approximation of $\delta\Phi(n_0+1, n, \omega)$ (20) we obtain:

$$e^{\Phi(n_0+1, n, \omega)} \sim e^{\Phi^{(sp)}(n_0+1, n, \omega)} [1 + \delta\Phi(n_0+1, n, \omega)].$$

However, while $\Phi(n_0 + 1, n, \omega)$ and $\Phi^{(sp)}(n_0 + 1, n, \omega)$ are normalized potentials, i.e., the log of a conditional probability, the first order approximation of $e^{\Phi^{(sp)}(n_0+1, n, \omega)} [1 + \delta\Phi(n_0 + 1, n, \omega)]$ is not. Normalization is obtained formally by introducing the partition function:

$$Z[\omega_{-\infty}^{n_0}] = \sum_{\omega_{n_0+1}^n} e^{\Phi^{(sp)}(n_0+1, n, \omega)} [1 + \delta\Phi(n_0 + 1, n, \omega)]$$

constrained by the past sequence $\omega_{-\infty}^{n_0}$, so that the quantity

$$\mathbb{P}^{(1)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \equiv \frac{e^{\Phi^{(sp)}(n_0+1, n, \omega)}}{Z[\omega_{-\infty}^{n_0}]} [1 + \delta\Phi(n_0 + 1, n, \omega)],$$

is the first order approximation of $\mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}]$.

Setting:

$$Z^{(sp)}[\omega_{-\infty}^{n_0}] = \sum_{\omega_{n_0+1}^n} e^{\Phi^{(sp)}(n_0+1, n, \omega)},$$

we may write, to first order:

$$\frac{1}{Z[\omega_{-\infty}^{n_0}]} = \frac{1}{Z^{(sp)}[\omega_{-\infty}^{n_0}]} \left[1 - \sum_{\omega_{n_0+1}^n} \frac{e^{\Phi^{(sp)}(n_0+1, n, \omega)}}{Z^{(sp)}[\omega_{-\infty}^{n_0}]} \delta\Phi(n_0 + 1, n, \omega) \right].$$

But, as $\Phi^{(sp)}$ is the log of a conditional probability, $Z^{(sp)}[\omega_{-\infty}^{n_0}] = 1$. So, finally, we obtain, to first order:

$$\mathbb{P}^{(1)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \sim \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \left[1 + \delta\Phi(n_0 + 1, n, \omega) - \mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) \mid \omega_{-\infty}^{n_0}] \right],$$

where:

$$\mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) \mid \omega_{-\infty}^{n_0}] = \sum_{\omega_{n_0+1}^n} \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \delta\Phi(n_0 + 1, n, \omega)$$

denotes the conditional expectation with respect to $\mu^{(sp)}$.

B Perturbed Average of Observables

We consider a time-dependent observable f with range R_f . We use here the notation R_f to differentiate the range of f from the range of the potential R , as these two range are not necessarily equal. We assume $t - t_0 > R_f$. We set $R_f = D_f + 1$. We have:

$$\mu[f(t, \cdot)] = \sum_{\omega_{n-D_f}^n} f(t, \omega_{n-D_f}^n) \mu[\omega_{n-D_f}^n] = \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \mu[\omega_{n_0+1}^n].$$

The last equality comes from the fact that $f(t, \omega)$ is independent of $\omega_{-\infty}^{n_0}$, thus using (13):

$$\begin{aligned}\mu[f(t, \cdot)] &= \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} \mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mu(d\omega) \\ &= \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} \mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega),\end{aligned}$$

where the last equation holds because, on $\mathcal{A}_{-\infty}^{n_0}$, $\mu = \mu^{(sp)}$ (stimuli is switched off at time $t_0 \leq n_0$ and the effects on spikes are manifested from time $n_0 + 1$ on).

Thus, replacing $\mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}]$ by $\mathbb{P}^{(1)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}]$, we obtain, up to first order:

$$\begin{aligned}\mu[f(t, \cdot)] &\sim \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) \\ &+ \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \delta\Phi(n_0 + 1, n, \omega) \mu^{(sp)}(d\omega) \\ &- \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) \mid \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega).\end{aligned}$$

The first term is $\mu^{(sp)}[f(t, \cdot)]$ from (13),(14). The second term is:

$$\sum_{\omega_{n_0+1}^n} \int_{\mathcal{A}_{-\infty}^{n_0}} f(t, \omega) \delta\Phi(n_0+1, n, \omega) \mathbb{P}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) = \mu^{(sp)}[f(t, \cdot) \delta\Phi(n_0 + 1, n, \cdot)]$$

from the compatibility conditions of Gibbs measures, and because by assumption ($n - n_0 > R_f$), $f(t, \omega)$ does not depend on $\omega_{-\infty}^{n_0}$.

For the third term:

$$\begin{aligned}\sum_{\omega_{n_0+1}^n} \int_{\mathcal{A}_{-\infty}^{n_0}} f(t, \omega_{n-D_f}^n) \mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) \mid \omega_{-\infty}^{n_0}] \mathbb{P}^{(sp)}[\omega_{n_0+1}^n \mid \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) \\ = \mu^{(sp)}\left[f(t, \cdot) \mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) \mid \omega_{-\infty}^{n_0}]\right].\end{aligned}$$

But, by assumption $f(t, \omega_{n-D_f}^n)$ does not depend on $\omega_{-\infty}^{n_0}$ ($n - n_0 > D_f$), whereas by definition of the conditional expectation $\mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) \mid \omega_{-\infty}^{n_0}]$ is the projection on the sigma-algebra $\mathcal{F}_{-\infty}^{n_0}$. As a consequence we have:

$$\begin{aligned}\mu^{(sp)}\left[f(t, \cdot) \mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) \mid \omega_{-\infty}^{n_0}]\right] &= \mu^{(sp)}[f(t, \cdot)] \mu^{(sp)}\left[\mathbb{E}^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) \mid \omega_{-\infty}^{n_0}]\right] \\ &= \mu^{(sp)}[f(t, \cdot)] \mu^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot)].\end{aligned}$$

Therefore,

$$\mu[f(t, \cdot)] = \mu^{(sp)}[f(t, \cdot)] + \mu^{(sp)}[f(t, \cdot) \delta\Phi(n_0 + 1, n, \cdot)] - \mu^{(sp)}[f(t, \cdot)] \mu^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot)]$$

Finally,

$$\begin{aligned} \delta^{(1)}\mu[f(t, \cdot)] &= \mathcal{C}^{(sp)}[f(t, \cdot), \delta\Phi(n_0 + 1, n, \cdot)] \\ &= \sum_{r=n_0+1}^{n=[t]} \mathcal{C}^{(sp)}[f(t, \cdot), \delta\phi(r, \cdot)]. \end{aligned}$$

which is equation (21) in the paper.

Acknowledgments. RC was supported by an ERC advanced grant “Bridges”. BC was supported by Renvision # 600847 and partly supported by the ECOS-Sud program (grant C13E06).

References

- [1] L. Abbott and P. Dayan. The effect of correlated variability on the accuracy of a population code. *Neural Comput*, 11:91–101, 1999.
- [2] J. Alonso, W. Usrey, and W. Reid. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, (383):815–819, 1996.
- [3] V. Baladi. Linear response, or else. *Proceedings ICM Seoul*, 3:525–545, 2014.
- [4] I. Brivanlou, D. Warland, and M. Meister. Mechanisms of concerted firing among retinal ganglion cells. *Neuron*, 20(3):527–539, 1998.
- [5] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11:1621–1671, 1999.
- [6] B. Cessac. A discrete time neural network model with spiking neurons ii. dynamics with noise. *J. Math. Biol.*, 62:863–900, 2011.
- [7] B. Cessac. Statistics of spike trains in conductance-based neural networks: Rigorous results. *Journal of Mathematical Neuroscience*, 1(8), 2011.
- [8] B. Cessac and R. Cofre. Spike train statistics and Gibbs distributions. *J. Physiol. Paris*, 107:360–368, 2013.
- [9] R. Cofre and B. Cessac. Dynamics and spike trains statistics in conductance-based integrate-and-fire neural networks with chemical and electric synapses. *Chaos, Solitons and Fractals*, 50(8):13–31, 2013.

- [10] R. Cofre and B. Cessac. Exact computation of the maximum-entropy potential of spiking neural-network models. *Phys. Rev. E*, 89(052117), 2014.
- [11] L. Croner, K. Purpura, and E. Kaplan. Response variability in retinal ganglion cells of primates. *PNAS.*, 90:8128–8130, 1993.
- [12] P. Dayan and L.F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.
- [13] R. deCharms and M. Merzenich. Primary cortical representation of sounds by the coordination of action potentials. *Nature*, (381):610–613, 1996.
- [14] B. Ermentrout and D. Terman. *Mathematical Foundations of Neuroscience*. Springer, 1st edition, July 2010.
- [15] R. Fernandez and A. Galves. Markov approximations of chains of infinite order. *Bull. Braz. Math. Soc. (N.S.)*, 33(3):295–306, 2002.
- [16] R. Fernandez and G. Maillard. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.*, 118(3-4):555–588, 2005.
- [17] E. Ferrea, A Maccione, L Medrihan, T Nieu, D Ghezzi, P Baldelli, F Benfenati, and L. Berdondini. Large-scale, high-resolution electrophysiological imaging of field potentials in brain slices with microelectronic multielectrode arrays. *Frontiers in Neural Circuits.*, 6(80), 2012.
- [18] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.
- [19] M. Helmstaedter, K. Briggman, S Turaga, V. Jain, and H. Seung and W. Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500:168–174, 2013.
- [20] A.L Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerves. *J. Physiol. (Lond.)*, 117:500–544, 1952.
- [21] E.M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, MA, 2007.
- [22] R. Kubo. Statistical-mechanical theory of irreversible processes. *J. Phys. Soc. (Japan)*, 12:570–586, 1957.
- [23] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(1):255–284, 1966.
- [24] L. Lapicque. Recherches quantitatives sur l’excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.*, 9:620–635, 1907.

- [25] B. Lindner, B. Doiron, and A Longtin. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys Rev E.*, 72, 2005.
- [26] V. Lucarini, T. Kuna, J. Wouters, and D. Faranda. Relevance of sampling schemes in light of Ruelle’s linear response theory. *Nonlinearity*, 25(5):1311, 2012.
- [27] R. Marc, B. Jones and C. Watt, J. Anderson, and C. Sigulinsky. Retinal connectomics: Towards complete, accurate networks. *Prog. Retinal Eye Research.*, 37:114–162, 2013.
- [28] S. Nirenberg and P. Latham. Decoding neuronal spike trains: how important are correlations. *Proceeding of the Natural Academy of Science*, 100(12):7348–7353, 2003.
- [29] S. Ostojic, N. Brunel, and V. Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains . *The Journal of Neuroscience*, 33(29):10234–10253, 2009.
- [30] L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Comput. Neural Syst.*, 15(04):243–262, November 2004.
- [31] V. Pernice, B. Staude, S. Cardanobile S, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Comput Biol.*, 2011.
- [32] J. Pillow, L. Paninski, V. Uzzell, E. Simoncelli, and E.J. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci*, 25:11003–11013, 2005.
- [33] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes, Exploring the Neural Code*. The M.I.T. Press, 1996.
- [34] M. Rudolph and A. Destexhe. Analytical integrate and fire neuron models with conductance-based dynamics for event driven simulation strategies. *Neural Computation*, 18:2146–2210, 2006.
- [35] D. Ruelle. Differentiation of SRB states. *Commun. Math. Phys.*, 187:227–241, 1997.
- [36] D. Ruelle. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Statist. Phys.*, 95:393–468, 1999.
- [37] D. Ruelle. Differentiation of SRB states for hyperbolic flows. *Ergod. Th. & Dynam. Sys*, 28:613–631, 2008.
- [38] E. Schneidman, M.J. Berry II, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.
- [39] M. Shadlen and W. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci*, 18(10):3870–3896, 1998.

- [40] R. K Snider, J. Kabara, B. Roig, and A. Bonds. Burst firing and modulation of functional connectivity in cat striate cortex. *J. Physiol.*, 80(550):730–744, 1998.
- [41] H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir. Population coding in neuronal systems with correlated noise. *Phys Rev E*, 64, 2001.
- [42] G. Tkačik, O. Marre, T. Mora, D. Amodei, M.J. Berry II, and W. Bialek. The simplest maximum entropy model for collective behavior in a neural network. *J Stat Mech*, page P03011, 2013.
- [43] J. Trousdale, Y. Hu, E. Shea-Brown, and K. Josic. Impact of network structure and cellular response on spike time correlations. *PLoS Comput Biol.*, 8, 2012.
- [44] M. Tsodyks, T. Kenet, A. Grinvald, , and A. Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286(5446):1943–1946, 1999.
- [45] J.C Vasquez, A. Palacios, O. Marre, M.J Berry II, and B. Cessac. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris*, 106(3-4):120–127, May 2012.